

# IML\_project

Heikki Nenonen

2022-11-23

## Todo

- dummy classifier
- ~~class4 -> event/nonevent, week1-exe?~~
- ~~drop partlybad, pelkkää FALSEa~~
- varianssit mukana/ei mukana? ei one hot -> yksinkertaistaa liikaa ja tarkoitettu kategoriseen dataan
- date? paljon informaatiota, mutta halutaanko muuttujaksi <- opeta 2000-2008, testaa 2009-2011 / kysy slack test\_hidden ei - - date, jätetäänkö pois?
- ~~train, test, cv-10?~~
- ~~itse logistiregression, week2-exe1 <- lasso/ridge~~
- ~~accuracy, perplexity, week2-exe1~~
- accuracy of our accuracy? <- malli train+test, vähän parempi kuin pelkkä train?
- class4 -> nonevent/1a/1b/II/
- googlaa mahdollisia malleja

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

```
#npf_test = pd.read_csv("initial_data/npf_test_hidden.csv")
npf_train = pd.read_csv("initial_data/npf_train.csv")
```

```
npf_train_test = npf_train.set_index("date")
npf_train_test = npf_train_test.drop(['id', 'partlybad'], axis=1)
```

```
class2 = np.array(["nonevent", "event"])
class2 = class2[(npf_train_test["class4"]!="nonevent").astype(int)]
#class2 = class2.apply(lambda x: 1 if "event" else 0)
npf_train_test.insert(loc=0, column="class2", value=class2)
npf_train_test["class2"].replace(["event", "nonevent"],[1,0], inplace=True)

npf_train_test
```

```
##           class2    class4  C02168.mean  C02168.std  C02336.mean  \
## date
## 2000-01-17      1         Ib    368.771711    0.310309    368.665658
```

## 2000-02-28	0	nonevent	378.197295	1.001493	378.083089
## 2000-03-24	1	Ib	373.043158	0.749865	372.930066
## 2000-03-30	1	II	375.643019	0.558629	375.549430
## 2000-04-04	0	nonevent	377.661030	0.408421	377.609576
## 2000-04-07	1	Ia	373.869464	0.655604	373.788580
## 2000-04-09	1	Ib	376.006588	1.109789	375.888889
## 2000-04-14	0	nonevent	374.068239	1.257096	374.042330
## 2000-04-20	0	nonevent	380.791148	1.872026	380.541141
## 2000-05-18	1	Ib	372.549952	4.454107	372.658235
## 2000-05-29	0	nonevent	375.497689	2.981294	375.151244
## 2000-06-02	0	nonevent	371.449301	4.533294	371.588297
## 2000-06-11	1	Ib	364.304231	3.925476	364.366111
## 2000-06-14	1	Ib	369.195304	1.022733	369.277363
## 2000-06-15	1	Ib	368.569447	1.907326	368.706144
## 2000-06-17	1	Ib	369.368178	2.112800	369.605805
## 2000-06-19	0	nonevent	369.409661	4.350994	369.587712
## 2000-06-21	1	II	369.673348	5.130404	369.566652
## 2000-06-22	0	nonevent	367.847076	7.417310	367.422627
## 2000-06-23	0	nonevent	374.410085	6.564796	373.976737
## 2000-07-04	0	nonevent	362.514311	7.334843	362.209307
## 2000-07-05	1	II	367.323656	7.562947	366.469737
## 2000-07-10	1	II	365.635240	10.538028	364.159693
## 2000-08-12	0	nonevent	373.794000	13.312325	373.526615
## 2000-09-21	1	Ia	365.816599	5.832111	366.087891
## 2001-05-16	1	Ib	370.710816	2.535541	370.931149
## 2001-05-17	0	nonevent	377.198972	2.091652	377.205374
## 2001-05-31	1	II	370.090881	0.753314	370.320705
## 2001-06-02	1	II	368.530393	3.543669	368.640131
## 2001-06-03	1	II	368.388341	5.323841	367.801834
## ...	...	...	...	...	...
## 2010-12-24	0	nonevent	406.590000	0.814045	406.145957
## 2010-12-29	0	nonevent	401.545000	0.315325	401.509583
## 2010-12-31	0	nonevent	405.562857	1.392169	405.454792
## 2011-01-05	0	nonevent	421.511176	0.335486	421.057843
## 2011-01-11	0	nonevent	397.321111	0.362916	397.373889
## 2011-01-14	0	nonevent	398.854464	0.337250	398.817500
## 2011-01-19	0	nonevent	403.834167	0.636059	403.743000
## 2011-01-26	0	nonevent	398.756615	0.451941	398.484308
## 2011-01-31	0	nonevent	397.970857	0.380502	397.928143
## 2011-02-26	0	nonevent	402.118526	1.696298	402.097158
## 2011-02-28	0	nonevent	404.520206	1.400434	404.454124
## 2011-05-29	0	nonevent	391.467363	2.957007	391.562527
## 2011-06-03	1	II	392.115269	2.170120	392.313495
## 2011-06-16	0	nonevent	379.813602	3.255927	380.343710
## 2011-06-19	1	II	387.945654	8.594281	387.195208
## 2011-06-28	1	II	381.683141	5.724127	381.709316
## 2011-07-02	0	nonevent	395.612897	15.629503	395.551869
## 2011-07-07	0	nonevent	378.837063	5.947588	378.781702
## 2011-07-15	0	nonevent	379.962912	9.685015	380.476374
## 2011-07-23	0	nonevent	388.033466	13.690201	387.962273
## 2011-07-26	1	II	380.884885	5.084053	380.694566
## 2011-08-01	1	II	377.558690	6.721647	377.967679
## 2011-08-09	0	nonevent	382.693688	6.599782	382.935652
## 2011-08-10	1	II	378.499375	9.021312	378.433438

##	2011-08-12	0	nonevent	377.215823	8.565593	377.672025
##	2011-08-16	0	nonevent	381.016623	4.411571	381.030844
##	2011-08-19	0	nonevent	383.698146	8.418351	384.052632
##	2011-08-21	0	nonevent	379.279128	12.045785	379.363087
##	2011-08-22	0	nonevent	384.443758	6.413297	384.364392
##	2011-08-27	0	nonevent	382.230839	6.407469	382.448601
##						
##		C02336.std	C0242.mean	C0242.std	C02504.mean	C02504.std \
##	date					
##	2000-01-17	0.305127	369.371184	0.333606	368.616053	0.302364
##	2000-02-28	1.025472	378.671311	1.017208	378.029508	1.022258
##	2000-03-24	0.728411	373.569735	0.835240	372.869868	0.692828
##	2000-03-30	0.540964	376.045849	0.580730	375.506013	0.529970
##	2000-04-04	0.423963	378.117134	0.417243	377.582242	0.424460
##	2000-04-07	0.649312	374.356310	0.629567	373.737083	0.633824
##	2000-04-09	1.052157	376.779883	1.403241	375.806257	0.960352
##	2000-04-14	1.214805	374.725480	1.374858	374.048523	1.184708
##	2000-04-20	1.571144	381.550380	2.291529	380.341522	1.420260
##	2000-05-18	4.424929	373.353381	4.678041	372.557286	4.159152
##	2000-05-29	2.169545	377.234646	3.412408	374.657467	1.592309
##	2000-06-02	4.505640	372.143509	4.897977	371.564323	4.472209
##	2000-06-11	3.728793	364.858333	4.185291	364.399744	3.617832
##	2000-06-14	0.924952	369.519834	1.362706	369.331978	0.894767
##	2000-06-15	1.769573	368.879787	2.199349	368.707745	1.656756
##	2000-06-17	1.965751	369.798894	2.876070	369.682797	1.867745
##	2000-06-19	4.216838	369.764280	4.536101	369.680890	4.134760
##	2000-06-21	5.014771	371.598584	6.286716	369.344872	4.771968
##	2000-06-22	6.535788	368.767627	7.850907	366.909322	5.515720
##	2000-06-23	5.720135	376.178596	9.051943	373.565404	5.155086
##	2000-07-04	6.631164	367.172244	12.074101	362.127696	6.363997
##	2000-07-05	6.769661	371.164361	12.407601	365.856987	6.011728
##	2000-07-10	7.877706	370.181542	16.066706	363.739515	7.034672
##	2000-08-12	12.715117	378.705722	17.344124	373.005179	11.716828
##	2000-09-21	5.731988	367.811438	6.439866	365.910270	5.221409
##	2001-05-16	2.524615	371.540270	4.139887	371.025405	2.550851
##	2001-05-17	2.228677	378.211355	3.036278	377.075869	2.308042
##	2001-05-31	0.799263	370.732863	2.140255	370.389207	0.785920
##	2001-06-02	3.400136	369.279561	4.245446	368.581228	3.137988
##	2001-06-03	4.408715	369.788478	7.029583	367.514934	4.007155
##	...	...	...	...	...	...
##	2010-12-24	0.984620	408.258261	2.648627	404.713191	1.178651
##	2010-12-29	0.333026	401.701875	0.378388	401.473750	0.299422
##	2010-12-31	1.331376	405.702708	1.316179	405.467292	1.325135
##	2011-01-05	0.511810	422.627800	1.079785	419.924510	0.818526
##	2011-01-11	0.358984	397.506792	0.336365	397.287963	0.340352
##	2011-01-14	0.352349	399.018246	0.387128	398.829464	0.335773
##	2011-01-19	0.568520	404.143051	0.837740	403.662542	0.541730
##	2011-01-26	0.541081	399.173333	0.859879	397.867121	1.095542
##	2011-01-31	0.409812	398.085072	0.368840	397.852286	0.420218
##	2011-02-26	1.709017	402.191809	1.708389	402.098211	1.763596
##	2011-02-28	1.331791	404.692917	1.511427	404.407010	1.260597
##	2011-05-29	2.915075	391.762692	3.133400	391.634754	2.862149
##	2011-06-03	2.094373	392.441568	2.688156	392.376162	1.943292
##	2011-06-16	2.958410	380.621622	4.083898	380.451290	2.557782

##	2011-06-19	6.395089	390.564398	12.503105	386.868542	5.561891
##	2011-06-28	5.354609	385.466526	10.259716	381.632670	5.014191
##	2011-07-02	14.612275	399.223491	19.976620	394.189439	12.953841
##	2011-07-07	5.560539	384.863147	13.079405	378.519291	5.106424
##	2011-07-15	9.170482	381.963132	10.555950	380.320437	8.547546
##	2011-07-23	12.878477	390.622971	15.321954	387.520114	11.730093
##	2011-07-26	3.964685	387.263103	11.950810	380.600347	3.363024
##	2011-08-01	6.693381	380.421916	8.069070	378.130888	6.529449
##	2011-08-09	6.366806	384.725375	8.754265	382.986894	6.187787
##	2011-08-10	7.588108	381.150126	12.054702	378.059625	5.942777
##	2011-08-12	7.475200	385.629108	20.851661	377.657278	6.605750
##	2011-08-16	4.062115	382.901742	5.061991	381.072078	3.840630
##	2011-08-19	8.130389	386.436800	9.901536	384.105592	7.952936
##	2011-08-21	11.533353	381.268591	14.043508	379.163200	10.932955
##	2011-08-22	5.781036	386.384762	7.663215	384.186892	5.506259
##	2011-08-27	6.242533	384.061748	7.748701	382.502587	6.174042
##						
##	...	T672.mean	T672.std	T84.mean	T84.std	UV_A.mean \
##	date	...				
##	2000-01-17	...	-1.476016	0.250952	-0.899393	0.271648 2.492491
##	2000-02-28	...	1.479493	0.313460	2.033191	0.271090 0.295937
##	2000-03-24	...	-0.270844	2.037111	0.780422	2.334741 14.434789
##	2000-03-30	...	5.622308	2.509758	6.533544	2.695277 16.077513
##	2000-04-04	...	1.140121	1.438071	2.133774	1.614122 9.710422
##	2000-04-07	...	2.079822	1.610976	3.181058	1.929516 16.688892
##	2000-04-09	...	3.499965	2.831278	4.677937	3.161601 17.456796
##	2000-04-14	...	1.421619	0.934736	2.009219	0.929537 4.279844
##	2000-04-20	...	10.424477	2.736685	10.534913	3.286528 9.526991
##	2000-05-18	...	15.256143	3.539267	16.335799	3.854709 18.891266
##	2000-05-29	...	11.128137	1.544570	11.010887	2.386529 5.693659
##	2000-06-02	...	9.186049	2.379577	10.035131	2.743211 11.902437
##	2000-06-11	...	18.538462	3.253127	19.427171	3.492863 17.506348
##	2000-06-14	...	13.019691	1.076942	13.922238	1.348261 20.666901
##	2000-06-15	...	9.746190	1.547373	10.652653	1.841738 18.457385
##	2000-06-17	...	6.835437	1.615293	7.569975	1.785952 11.966479
##	2000-06-19	...	10.953525	2.643322	11.825136	2.834392 12.299269
##	2000-06-21	...	18.159408	3.977961	18.965480	4.368491 21.563634
##	2000-06-22	...	21.536585	2.903180	22.158898	3.859714 18.617744
##	2000-06-23	...	16.709958	1.070875	17.285827	1.679039 17.419743
##	2000-07-04	...	17.874488	2.204949	18.366305	2.727982 16.301268
##	2000-07-05	...	17.284309	2.630577	17.806390	3.311183 16.403855
##	2000-07-10	...	13.860158	2.831995	14.169851	3.261185 19.498308
##	2000-08-12	...	12.645359	0.899486	13.290113	1.115743 7.373144
##	2000-09-21	...	8.542367	2.434650	9.066939	2.758029 12.636482
##	2001-05-16	...	12.146676	3.800598	12.966092	4.175156 21.084825
##	2001-05-17	...	8.851141	0.712621	9.395809	0.806294 5.468890
##	2001-05-31	...	9.059224	2.258354	9.911489	2.826368 20.067543
##	2001-06-02	...	11.348117	2.483802	12.316217	3.135771 21.142716
##	2001-06-03	...	10.964738	1.772399	11.623595	2.410785 11.695691
##	...	...	...	...	...	...
##	2010-12-24	...	-23.223099	0.661812	-24.874583	0.352509 1.986904
##	2010-12-29	...	-13.125831	0.304615	-12.300815	0.284083 0.898580
##	2010-12-31	...	-10.236087	0.189649	-9.595124	0.214066 1.430363
##	2011-01-05	...	-13.102716	0.397214	-13.285313	0.424018 1.904961

##	2011-01-11	...	-1.482958	0.318266	-0.869775	0.309397	0.986014
##	2011-01-14	...	-9.180490	0.170793	-8.414387	0.165328	1.913300
##	2011-01-19	...	-2.435476	0.114793	-1.814087	0.193177	2.236889
##	2011-01-26	...	-13.375731	1.206013	-16.405920	1.579065	3.603953
##	2011-01-31	...	-5.513822	0.513898	-5.032933	0.741341	2.233162
##	2011-02-26	...	-3.501131	0.478360	-2.770616	0.375086	2.342078
##	2011-02-28	...	-8.075873	1.256866	-7.260962	1.720847	8.374139
##	2011-05-29	...	7.759662	0.940267	8.390213	1.083921	6.334116
##	2011-06-03	...	13.561424	2.315332	14.545100	2.496177	19.766555
##	2011-06-16	...	15.164987	2.336026	15.281898	3.168567	19.018731
##	2011-06-19	...	13.846229	1.908796	14.671475	2.278146	14.888953
##	2011-06-28	...	19.604545	3.702358	20.096451	3.919209	22.337789
##	2011-07-02	...	22.371866	1.765061	22.674164	2.717245	15.173941
##	2011-07-07	...	21.626106	2.471505	21.967780	3.296952	19.782024
##	2011-07-15	...	15.561149	1.956759	16.159457	2.467483	11.921088
##	2011-07-23	...	22.733969	2.191116	23.527735	2.663090	17.161429
##	2011-07-26	...	21.090999	2.008273	21.263856	2.919077	18.588745
##	2011-08-01	...	12.421165	1.097733	13.260500	1.828132	15.369565
##	2011-08-09	...	13.911922	1.070119	14.541529	1.642235	12.015140
##	2011-08-10	...	12.419919	1.154197	12.957452	1.791148	13.148561
##	2011-08-12	...	13.846561	2.066032	14.187257	2.621028	16.918405
##	2011-08-16	...	16.319361	1.089563	17.268471	1.308728	8.688739
##	2011-08-19	...	15.820245	2.868514	16.352391	3.055122	18.475262
##	2011-08-21	...	14.421092	1.696894	15.094141	1.745526	8.099394
##	2011-08-22	...	15.049728	0.969001	16.076217	1.197644	11.665070
##	2011-08-27	...	21.040539	2.466867	21.946135	2.604069	16.670267
##							
##		UV_A.std	UV_B.mean	UV_B.std	CS.mean	CS.std	
##	date						
##	2000-01-17	1.310880	0.031587	0.018122	0.000243	0.000035	
##	2000-02-28	0.177836	0.005140	0.003552	0.003658	0.000940	
##	2000-03-24	8.627312	0.353743	0.272472	0.000591	0.000191	
##	2000-03-30	9.984686	0.568242	0.451830	0.002493	0.000466	
##	2000-04-04	7.054069	0.339135	0.291457	0.004715	0.000679	
##	2000-04-07	10.504951	0.561251	0.451130	0.000851	0.000244	
##	2000-04-09	10.967471	0.716453	0.572409	0.002083	0.000203	
##	2000-04-14	2.425409	0.146308	0.106017	0.002650	0.000891	
##	2000-04-20	8.514172	0.378214	0.406251	0.010655	0.002020	
##	2000-05-18	13.617161	0.895416	0.789068	0.007836	0.000906	
##	2000-05-29	3.436786	0.218611	0.153834	0.006468	0.001546	
##	2000-06-02	11.783641	0.500939	0.566143	0.005309	0.001051	
##	2000-06-11	13.955493	0.856997	0.828865	0.010701	0.003549	
##	2000-06-14	15.602926	0.906659	0.775303	0.001981	0.000329	
##	2000-06-15	14.970213	0.777403	0.780244	0.001316	0.000335	
##	2000-06-17	8.546585	0.473783	0.440850	0.001212	0.000331	
##	2000-06-19	10.471825	0.598677	0.588339	0.004936	0.001156	
##	2000-06-21	15.202560	1.031292	0.918180	0.004106	0.000862	
##	2000-06-22	14.027402	0.910155	0.857661	0.009253	0.001208	
##	2000-06-23	16.830501	0.959425	1.055615	0.007524	0.004676	
##	2000-07-04	12.283423	0.742604	0.717554	0.004911	0.000924	
##	2000-07-05	13.710872	0.811088	0.782701	0.003297	0.000545	
##	2000-07-10	15.942172	0.867749	0.821457	0.002602	0.000598	
##	2000-08-12	4.709601	0.315726	0.258542	0.006143	0.001479	
##	2000-09-21	7.657292	0.606501	0.463601	0.002336	0.000860	

```

## 2001-05-16 14.455005 0.815426 0.707169 0.002828 0.000716
## 2001-05-17 3.227746 0.167661 0.125177 0.007673 0.001082
## 2001-05-31 15.132695 0.749871 0.702205 0.001850 0.000348
## 2001-06-02 15.178141 0.805438 0.726420 0.003176 0.000517
## 2001-06-03 9.173489 0.425933 0.400514 0.005079 0.003658
## ... ... ... ... ...
## 2010-12-24 0.806453 0.050125 0.021531 0.002743 0.000284
## 2010-12-29 0.392891 0.024831 0.011980 0.002181 0.000083
## 2010-12-31 0.694178 0.036220 0.018626 0.003507 0.000589
## 2011-01-05 0.833850 0.049601 0.023256 0.004290 0.000234
## 2011-01-11 0.608848 0.025028 0.016643 0.001554 0.000357
## 2011-01-14 0.977117 0.044266 0.024131 0.001127 0.000251
## 2011-01-19 1.199880 0.044399 0.028712 0.001453 0.000106
## 2011-01-26 1.792172 0.084675 0.044226 0.001926 0.000221
## 2011-01-31 1.027305 0.057590 0.027986 0.001955 0.000346
## 2011-02-26 1.213024 0.090870 0.054802 0.006140 0.001231
## 2011-02-28 4.754612 0.295970 0.196800 0.006240 0.000189
## 2011-05-29 5.185683 0.305230 0.310133 0.002941 0.000378
## 2011-06-03 14.770082 0.988528 0.872272 0.001721 0.000620
## 2011-06-16 13.889024 0.879685 0.753987 0.003260 0.000288
## 2011-06-19 13.891135 0.724499 0.741484 0.002978 0.000775
## 2011-06-28 15.794595 1.197271 1.042756 0.003555 0.001285
## 2011-07-02 12.658809 0.755829 0.758103 0.007876 0.001519
## 2011-07-07 14.406956 0.971931 0.837506 0.004739 0.000706
## 2011-07-15 9.532734 0.605303 0.539926 0.004614 0.000667
## 2011-07-23 12.328885 0.831893 0.721734 0.007750 0.001455
## 2011-07-26 13.239182 0.909833 0.781493 0.004650 0.000492
## 2011-08-01 10.989756 0.818469 0.698631 0.001909 0.000567
## 2011-08-09 7.597650 0.540294 0.402583 0.002385 0.000647
## 2011-08-10 8.464131 0.616689 0.484763 0.001068 0.000343
## 2011-08-12 11.929451 0.787474 0.661421 0.002091 0.000152
## 2011-08-16 8.320799 0.464422 0.496816 0.002423 0.000425
## 2011-08-19 12.838481 0.884404 0.726461 0.002476 0.000902
## 2011-08-21 5.851942 0.416961 0.363890 0.003484 0.000457
## 2011-08-22 10.062797 0.622831 0.595032 0.004782 0.001082
## 2011-08-27 11.216294 0.899029 0.722553 0.006956 0.000605
##
## [464 rows x 102 columns]

```

```

X = npf_train_test.drop(["class2", "class4"], axis=1)
y = npf_train_test["class2"]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, train_size=100, random_state=42, shuffle=True, stratify=npf_train_test["class2"]
)

def loss(X_tr, y_tr, X_te, y_te, m):
    return mean_squared_error(y_te, m.fit(X_tr, y_tr).predict(X_te), squared=False)

def accuracy(X_tr, y_tr, X_te, y_te, m):
    return accuracy_score(y_te, m.fit(X_tr, y_tr).predict(X_te))

perplexity = lambda p: np.exp(-np.mean(np.log(y_test*p + (1 - y_test) * (1 - p))))

```

```

models = [DummyClassifier(), LogisticRegression(), LogisticRegression(penalty="elasticnet", l1_ratio=0.5)]

res = pd.DataFrame(index=models)

# Loss on training data, for model trained on training data:
res["train_loss"] = [loss(X_train, y_train, X_train, y_train, m) for m in models]
# Los on test data, for model trained on training data:

## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning:
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)

res["test_loss"] = [loss(X_train, y_train, X_test, y_test, m) for m in models]
# Cross-validation loss:

## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning:
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)

res["cv_loss"] = [
    -cross_val_score(
        m, X_train, y_train, cv=10, scoring="neg_root_mean_squared_error"
    ).mean()
    for m in models
]

## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning:

```

```

## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_logistic.py:765: ConvergenceWarning
## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```







```

## STOP: TOTAL NO. of ITERATIONS EXCEEDS LIMIT.
##
## Increase the number of iterations (max_iter) or scale the data as shown in:
##     https://scikit-learn.org/stable/modules/preprocessing.html
## Please also refer to the documentation for alternative solver options:
##     https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
##     extra_warning_msg=_LOGISTIC_SOLVER_CONVERGENCE_MSG)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
## /home/pniko/.local/lib/python3.6/site-packages/sklearn/linear_model/_sag.py:329: ConvergenceWarning:
## "the coef_ did not converge", ConvergenceWarning)
res["test_perplexity1"] = [perplexity(m.predict_proba(X_test)[: ,1]) for m in models]
res

##                                     train_loss  test_loss  \
## DummyClassifier()                    0.707107   0.707107
## LogisticRegression()                 0.264575   0.416025
## LogisticRegression(l1_ratio=0.5, penalty='elast... 0.374166   0.435385
## LogisticRegression(C=1, solver='saga')          0.374166   0.435385
## LogisticRegression(penalty='none', solver='saga') 0.374166   0.435385
## LogisticRegression(C=1, penalty='l1', solver='s... 0.374166   0.435385
##
##                                     cv_loss  test_accuracy  \
## DummyClassifier()                    0.707107         0.500000
## LogisticRegression()                 0.393349         0.826923
## LogisticRegression(l1_ratio=0.5, penalty='elast... 0.370200         0.810440
## LogisticRegression(C=1, solver='saga')          0.370200         0.810440
## LogisticRegression(penalty='none', solver='saga') 0.370200         0.813187
## LogisticRegression(C=1, penalty='l1', solver='s... 0.370200         0.810440
##
##                                     test_perplexity1
## DummyClassifier()                    2.000000
## LogisticRegression()                 1.738969
## LogisticRegression(l1_ratio=0.5, penalty='elast... 1.511714
## LogisticRegression(C=1, solver='saga')          1.510755
## LogisticRegression(penalty='none', solver='saga') 1.510560
## LogisticRegression(C=1, penalty='l1', solver='s... 1.511521

```